

Helping the young engineering faculty to understand the peer-review process

Luis A. Godoy, Ph.D.

Professor and Associate Director for Research, Department of Civil Engineering and Surveying
University of Puerto Rico at Mayagüez
Mayagüez, PR 00681-9041, Puerto Rico.

Abstract

This paper reports on a controlled experiment carried out to investigate the differences between novices and experts during the review of papers for scientific journals. A total of 17 young faculty members with Ph. D. were given instruction on the peer review process, and then had to review a short paper, for which the actual reviews were available. The quality of the reviews was assessed using criteria developed and widely accepted in Medical journals. The evidence obtained from this study helps to identify the critical areas in which novices fail to perform a good review, and then to develop strategies to overcome such limitations.

Keywords

Engineering education, Faculty development, Journal Publications, Peer-review system.

1. Introduction

Traditionally, the new engineering faculty at universities does not have any source of information or training about their new job, and have to learn the office by themselves. This approach may change in the next few years, because some universities recognize the need to implement some formal mentoring to make things easier to the newcomer and to follow her progress. Strategies for the success of new faculty have been proposed by several authors [6]. This paper addresses the topic of serving as reviewers for scientific journals, which is a task soon required from engineering faculty and for which, again, there is no training.

Peer-review is a process of quality control by peers, according to standards. The system of journal publication heavily relies on the peer-review system, in which each manuscript is reviewed by two or three researchers who remain anonymous to the author. The aims of this process are twofold: To help the editor decide whether an article should be published or not (based on quality and relevance to the journal), and to improve the article before publication.

Although reviewers are people trained to do research, they have no training on how to perform a review. Large journals use many reviewers twice or three times a year; however, they offer no training to their reviewers. Furthermore, journal editors choose reviewers based on the ability to do research, without an evaluation of their ability to criticize a paper. And to complete the circle of lack of assessment, editors do

not evaluate the actual reviewers work in order to identify who does a poor job and who has provided a substantiated report on the merits of a manuscript.

Some of the most sophisticated methods of assessing reviewers are ratings done by the editors on individual reviews, using a subjective scale between 1 and 5 [2, 8, 3]. A second procedure to rate reviewers is reported by Baxt et al. [1], in which “reviewers were sent a fictitious manuscript, which contained 23 deliberate flaws and was masked to its true purpose, to review”. A third action has been to train reviewers, in which “the reviewer received a copy of the other reviews of the same manuscript” [3].

As mentioned above, the main efforts in this direction have been made by journal editors (especially in Medicine), who face the problem of low quality reviews and misconduct in their daily editorial activity. But there is no reason why this problem should be in the hands of journal editors, and perhaps it would be an important part in the training of young faculty. Programs like NSF-AGEP (Alliance for Graduate Education and the Professoriate) or specific university programs, aimed at improving young faculty publications [4], deal with this problem in different ways.

This paper investigates the quality of reviews performed by novices, and compares them with reviews on the same manuscript made by experts. In Section 2, the main assumptions of the peer-review system for journal papers are stated. Section 3 reports on an experiment done with 17 young researchers, all with Ph. D., and already faculty members. Most were assistant professors and voluntarily enrolled in an initiative to strengthen their publication skills. The criteria for quality assessment are in Section 4, together with the analysis of the reviews of novices and experts following these criteria. Finally, some conclusions are drawn in Section 5.

2. Main Assumptions of the Peer Review System

The peer-review process involves authors, editors, and reviewers. The part related to the reviewers works under the following assumptions:

- (a) The evaluation of a paper by a small number of reviewers is sufficient to evaluate the merits of a scientific manuscript submitted to a journal. Most engineering journals use either two or three reviewers, some use just one. In other fields the number of reviewers is much larger, such as in anthropology and social science journals.
- (b) A reviewer knows the state of the art in the field of the manuscript. One of the reasons why the reviewer has been chosen by the editor is because she was identified as an expert in the field, perhaps because of previous publications in this or other journals, or because she was recommended by other experts. A qualified reviewer eliminates/reduces the possibilities of plagiarism, fraud, duplicate publications by an author or attempt to publish material that has already been published by others.
- (c) The reviewer understands in depth the manuscript. This means that she has dedicated sufficient time to the paper to understand the general aspects and also the details. Notice that if a reviewer is an expert in the field and has made a conscientious effort to study the manuscript, and even then she does not understand the manuscript, then there is a problem with the manuscript and not with the review process. A careful reviewer eliminates/reduces the possibility of data falsification and flaws in the analysis.
- (d) The reviewer evaluates the research reported in the manuscript, and not the author or her environment. Thus the outcome of the review should not depend on sex, status, race of the author, or the institution in which the researcher works. An honest reviewer eliminates/reduces the possibility of bias.
- (e) The specific scientific interests of the reviewer do not influence the outcome of the review. This means that the reviewer should not take into account how the manuscript affects (either positively or negatively) her own interest in the field, past, present or future work.
- (f) The reviewer does not use the findings of the manuscript for her own benefit or in her own work. And she does not share the manuscript with other researchers or students without the explicit previous consent of the editor.

Assumptions (b-f) are associated to ethical conducts of the reviewer. The reviewer should decline performing a review if she believes that has not followed the literature in the field for some time (to avoid violation of assumption b); cannot dedicate sufficient time to the review (assumption c); has personal situations which would be in conflict with an objective review, such as being relatives, members of the same team, had personal problems in the past, and others (assumption d); has conflict of interest with the work reported in the manuscript (assumption e).

Problems in the peer-review system are associated to failure to accomplish the assumptions in which the system is based. The study of Peters and Ceci [5] highlights the difficulties regarding assumptions (b-d). And there are numerous cases of fraud, data falsification, plagiarism, which could be avoided by a more careful respect for the assumptions (for example, the Soman and Felig case, reported in [6, pp. 169-170]). Of course, this does not summarize all the ethical problems in the peer-review system, but are only the most frequent ones in relation to the evaluation of a manuscript.

3. Experiment

To evaluate the differences between experts and novices when they review papers, we carried out a controlled experiment with a small number of young faculty from Civil, Mechanical, and Industrial Engineering. The study was done three times in different years, each time with a group of six participants who learned about the peer-review process in a workshop. The total number of participants was 17. All participants had already obtained their Ph. D., most in recent years, and most were tenure track assistant professors.

Because we wanted to give the same manuscript to all participants and they had different backgrounds, then we spent some time to choose the manuscript. The final choice was a real manuscript, written in English, which had gone through the peer-review process and for which we had the real reviews. The manuscript had been submitted to a prestigious journal as a technical note and was rejected. The contents were not so difficult, so they could be grasped by the participants provided they performed a careful reading. None complained about complexity of the manuscript or being too far from their fields of expertise.

Each participant was given the same short manuscript without a title or abstract. They studied the manuscript for one hour and wrote the missing parts. This was done to make sure that they understood the contents of the manuscript. These titles and abstracts were checked after the activity was completed and were good for the present purposes, so we understand that assumption (c) was satisfied. Then they had to perform a review of the manuscript. At the end of the review they produced a recommendation, which could be:

A : Accept the paper as it is.

AmiCh : Accept with minor changes.

AMaCh : Accept the manuscript provided major changes are introduced.

RejRes : Reject the manuscript as it is, and encourage the author to resubmit.

Rej : Reject.

Let us consider the relation with the assumptions of the peer-review process mentioned in the previous section. The double blind process in this case assured that there was no possible bias in the process, so that assumption (d) was satisfied. Since the participants were not linked to the field of the manuscript, then assumptions (e, f) were also satisfied. But clearly assumption (b) is violated, i.e. the participants were not experts in the field and do not master the state of the art.

4. Quality of the Reviews

To investigate the reviews, we used the Components of a Quality Review identified by Callaham et al. [3], which are:

C1. The reviewer identified and commented on major strengths and weaknesses of study design and methods.

C2. The reviewer commented accurately and productively on the quality of the author's interpretation of the data, including acknowledgement of the data's limitations.

C3. The reviewer commented on major strengths and weaknesses of the manuscript as a written communication, independent of the design, methods, results and interpretations of the results.

C4. The reviewer provided the author with useful suggestions for improvement of the manuscript.

C5. The reviewer's comments to the author were constructive and professional.

C6. The reviewer provided the editor with the proper context and perspective to make a decision about acceptance or revision of the manuscript.

The results are presented in Table 1. Each component was assigned a maximum of 1, and the scale used was [1, 0.5, 0]. The values were added to compute the score, with a maximum of 5. The fifth component C5 was considered in the analysis but not taken into account in the computation of the score because it is not associated to the main purpose of the present research. The highest score was 3.5, the lowest was 0.5, and the average value was 1.617.

It is interest to consider the individual components of quality. Most participants had a high score in components C3 and C5 (0.79 on average), very low score in C4 (only one participant made a significant recommendation to the authors), and intermediate to low values in the others (average 0.206 for C1 and C6; 0.353 for C2).

The final decision of the participants was 70.6% decided some form of acceptance (53% AmiCh; and 18% AMaCh), and 29.4% recommended rejection (24% RejRes; 6% Rej).

The table contains the quality components obtained from the reviews themselves for this manuscript (Exp1, Exp2, Exp3), which were available to the author of this paper. The three reviewers recommended rejection of the manuscript. The experts concentrated on the weakness of the manuscript, not on the strengths, and some did not have a constructive attitude. But they did a very good job in identifying the weakness of the paper and showing doubts that were not present in the report of the novices. The topics discussed by the experts in C1 and C2 were not identified by any of the 17 novices. The average score of the experts (3.66) was higher than the maximum score obtained by a participant (3.5). The experts did not question originality of the research, or lack of advantage with respect to existing knowledge, and their critique concentrated on the model assumed and on the lack of completeness of the study.

We also investigated the agreement between the recommendations of participants and experts. A value of 1 was given to the participants when they recommended Rej; 0.8 for RejRes; 0.4 for AMaCh; 0.2 for AmiCh; and 0.1 for A. Only one participant made the same recommendation as the experts, and none recommended acceptance as it is. As a combined measure of the score and the agreement with the experts, we multiplied the two values, and they are included in the last column in Table 1. The maximum possible value is 5, and one participant obtained 3.5, but the average value was 0.82, which was considered a low value.

5. Conclusions

The present study attempted to study the quality of reviews produced by a group of 17 novices, and to compare the results and recommendations with those of a group of 3 experts.

An analysis of the contents of the reviews shows that the novices concentrated on more superficial aspects, such as on the quality of the written communication, but failed to identify the strengths and

weakness of the manuscript. Furthermore, they were not able to make suggestions to the author regarding ways to improve the manuscript or the research. This produced a review which was not very useful in helping the editor to make a decision. The final recommendation of novices was towards accepting the manuscript; in contrast, the experts were unanimous in rejecting the paper in this particular case. The evidence obtained from this study helps to identify the critical areas in which novices fail to perform a good review, and then to develop strategies to overcome such limitations.

Our study has several limitations. There were a small number of participants involved in the study, and only one manuscript was considered. Furthermore, the assessment was performed using a scale developed by researchers for medical journals, which may have differences with those of engineering. Another limitation refers to the computation of combined factors to weight the quality of the review and the agreement with experts. Finally, the study cannot separate what part of the poor performance of novices is due to their lack of experience doing reviews and what part is associated to their lack of knowledge in the field.

Acknowledgements

The author thanks the participants of this study, who should remain anonymous. This work was supported by a grant provided by the University of Puerto Rico at Mayagüez.

References

- [1] Baxt W.G., Waekerle J.F., Tintinalli J.E., Knopp R.K., Callaham M.L., "Evaluation of the peer reviewer: performance of reviewers on a fictitious submission", *Acad. Emerg. Med.*, vol. 3 (1996) 504.
- [2] Callaham M.L., Baxt W., Waekerle J., Wears R., "The reliability of editor's subjective quality ratings of manuscript peer reviews", *Journal of the American Medical Association*, vol. 280 (1998) 229-231.
- [3] Callaham M.L., Knopp R.K., Gallagher E.J., "Effect of written feedback by editors on quality of reviews", *Journal of the American Medical Association*, vol. 287(21) (2002) 2781-2783.
- [4] Godoy L.A., Valeiras N., "An initiative to strengthen the peer-review publications in the young faculty", *ASCE Journal of Professional Issues in Engineering Education and Practice*, vol. 127(7) (2001) 116-121.
- [5] Harnad, S., *Peer Commentary on Peer Review*, Cambridge University Press, Cambridge, UK (1982).
- [6] McCuen, R.H. (Ed.), *The Elements of Academic Research*, American Society of Civil Engineers, New York (1996).
- [7] Narayanan, R.R., "Academic leadership strategies for engineering faculty", *Int. J. Engineering Education*, vol. 19(2) (2003) 241-251.
- [8] Van Rooyen S., Black N., Godlee F., "Development of the review quality instrument (RQI) for assessing peer review of manuscripts", *J. Clinical Epidemiology*, vol. 52 (1999) 625-629.

Table 1. Summary of results of the experiment.

| Participant | C 5 | C 1 | C 2 | C 3 | C 4 | C 6 | Score | Rec | Sc*Rec |
|-------------|------|-------|-------|------|------|-------|-------|--------|--------|
| P1 | 1 | 0 | 1 | 0.5 | 1 | 0.5 | 3 | AmiCh | 0.6 |
| P2 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | AmiCh | 0.2 |
| P3 | 1 | 0.5 | 1 | 1 | 0 | 1 | 3.5 | Rej | 3.5 |
| P4 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | AMaCh | 0.5 |
| P5 | 0.5 | 0 | 1 | 1 | 0 | 0.5 | 2.5 | RejRes | 2 |
| P6 | 0.5 | 0 | 0 | 1 | 0 | 0 | 1 | AmiCh | 0.2 |
| P7 | 0.5 | 0 | 0 | 1 | 0 | 0 | 1 | AmiCh | 0.2 |
| P8 | 1 | 0 | 1 | 1 | 0 | 0 | 2 | AMaCh | 1 |
| P9 | 0.5 | 0 | 0 | 1 | 0 | 0 | 1 | AmiCh | 0.2 |
| P10 | 1 | 0 | 0 | 1 | 0 | 0 | 1 | AmiCh | 0.2 |
| P11 | 0.5 | 0 | 0 | 1 | 0 | 0 | 1 | AmiCh | 0.2 |
| P12 | 1 | 1 | 1 | 0 | 0 | 1 | 3 | AMaCh | 1.5 |
| P13 | 1 | 1 | 0 | 0.5 | 0 | 0.5 | 2 | RejRes | 1.6 |
| P14 | 1 | 0 | 1 | 1 | 0 | 0 | 2 | AmiCh | 0.4 |
| P15 | 0.5 | 0 | 0 | 0.5 | 0 | 0 | 0.5 | RejRes | 0.4 |
| P16 | 0.5 | 1 | 0 | 0.5 | 0 | 0 | 1.5 | RejRes | 1.2 |
| P17 | 1 | 0 | 0 | 0.5 | 0 | 0 | 0.5 | AmiCh | 0.1 |
| Average | 0.79 | 0.206 | 0.353 | 0.79 | 0.06 | 0.206 | 1.617 | | 0.823 |
| Exp1 | 1 | 1 | 0 | 0 | 1 | 1 | 3 | RejRes | |
| Exp2 | 0.5 | 1 | 1 | 0 | 1 | 1 | 4 | Rej | |
| Exp3 | 0.5 | 1 | 1 | 0.5 | 0.5 | 1 | 4 | Rej | |
| | 0.67 | 1 | 0.67 | 0.17 | 0.83 | 1 | 3.67 | | |